caBIG
Architecture Workspace
caGRID Phase II

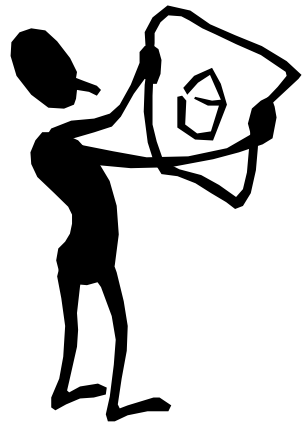
10/25/2004

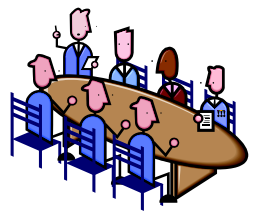
Agenda

1. ***Requirement Analysis – Preliminary results.***
   - Overview.
   - Key findings.
   - Use case examples.
   - caBIG Actors.
   - caBIG integrated use case diagram.
   - caBIG technical requirements.
2. ***Architecturally significant technologies – Panel***
   - W3C - Web services, Semantic Web. – Panthers team
   - GGF/OASIS - OGSA/OGSI/WS-RF. – OSU Team
   - Semantic Grid – SAIC Team
3. ***caBIG preliminary System Components***
   - Main components/layers – SAIC
4. ***Future tasks***

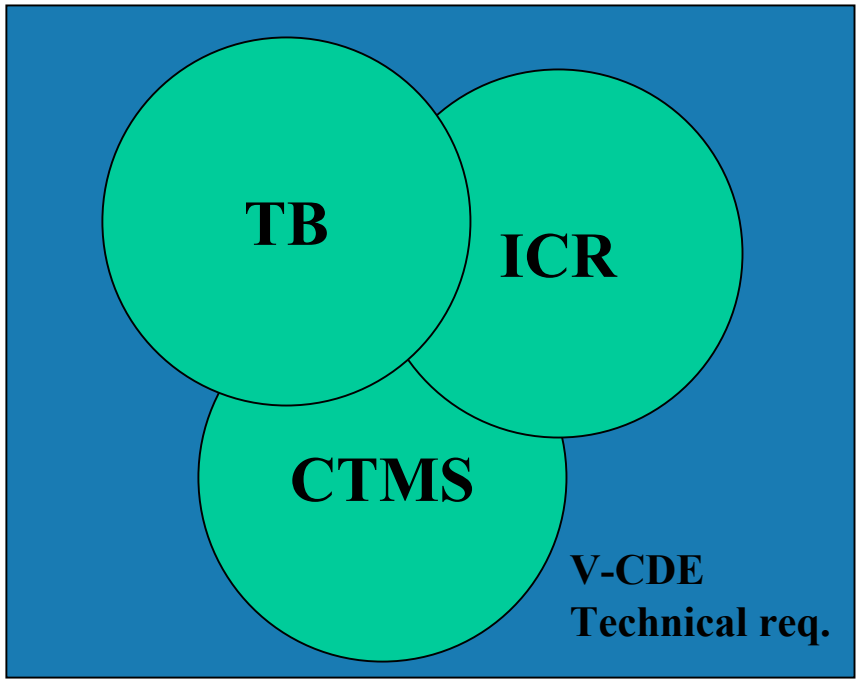
1. Requirement Analysis – Preliminary results.

-Overview.
-Key findings.
-Use case examples.
-Data provider.
-Analysis tool provider.
-Adopter.
-caBIG Actors.
-caBIG integrated use case diagram.
-caBIG technical requirements.





Overview – Requirement analysis



- Iteration process.
- Bottom-Up and Top down approach.
- Identify common patterns to define integration standards.
- Looking for collaboration use cases to support caBIG mission.

Use Case example 1:

- Use Case:
  - A researcher would like to perform a study to compare the type of cancer treated with a particular chemotherapeutic agent, involved in clinical trials, in which the study subjects lived longer than 2 years post-study to find gene expression patterns that might be predictive of a positive outcome. The researcher has expression data (Affy) for patients that lived less than two years.
- Query:
  - I want to collect all microarray data (Affy only) available from all cancer centers from patients with bladder or ovarian cancer that were part of any clinical trial protocol using cisplatin within the past five years. In addition, I want to know all available tissue samples, cancerous and non-cancerous (normal) tissue localized within 10mm of tumor site from this patient group such that I can perform Affy gene expression studies to include with previously performed studies that were identified by the query. Finally, I need all severe adverse events for the group of patients identified that had a severity rating of 3-4 and are likely linked to cisplatin administration.

Use case example 2

With the caBIG grid infrastructure in place, it should be possible to:

- An investigator logs on to the grid and searches for chips of interest in a caArray data source. After identifying a set of chips to analyze, he/she accesses a caGRID microarray analysis service (like VISDA, FGDP, Distance-Weighted-Discrimination, Gene Pattern, etc.) to analyze the data. Output from this service will be lists of "interesting" spots or probes, which may then be piped to annotation services (like caBIO, Function Express, GOMiner, Reactome Data, Cancer Molecular Pages, PIR, HapMap, PromoterDB, etc.)
- An investigator performs a proteomics experiment that is stored/tracked in the Proteomics LIMS and analyzes this data using RProteomics and a protein identification routine (like Mascot). Output from this analysis will yield lists of "interesting" spots corresponding to protein hits. Annotations for these proteins (using accession numbers from GenBank, SwissProt, Ensembl, etc. usually although search databases may use anything as identifiers) will be acquired from the annotation services listed in #1 above.

Use case example 3

Use Case 3: Setting Response Criteria

CHARACTERISTIC INFORMATION

Goal in Context: To set response criteria for the results returned by UniProt/PIR web service

Preconditions:

caGRID query client is ready to submit a query to UniProt/PIR web service.

caGRID query client provides an interface to set the criteria on the results returned by UniProt/PIR web service.

Success End Condition: The user sets response criteria

Failed End Condition: The user is unable to find the response options that cover the information needed.

Primary Actor: Researcher, Scientist

MAIN SUCCESS SCENARIO

The user sets the information content of the result returned by UniProt/PIR web service. The user may choose to use the information provided by default response(in UniProt XML format) , in which each protein record contains:

UniProt ID and accession number(s)

Protein name(s) and components

Gene name(s) and symbol(s)

Keywords

Scientific and common organism name

Software feature list

Preliminary results

	MSKCC	Wistar (L)	Wistar (J)	Washington	UPENN	Georgetown	UNC
Query caBIG resources	X	X	X	X	X	X	X
Update data sources		X					
Advertise data source		X	X	X	X	X	
Advertise anal. Tool		X	X				X
Discover caBIG resources		X	X	X		X	X
Notification of resource update		X	X		X	X	X
Query literature			X	X			
Annotation			X	X			
Analysis tools		X	X	X			X
ID Resolution	X				X		X
Data mapping		X		X			
Metadata (*)	X	X	X	X	X	X	X
Workflow analysis tools		X		X			
Quality control		X					
Security				X	X		
Performance							

* Includes version tracking, provenance, CDE

Adopter Data provider Tool provider

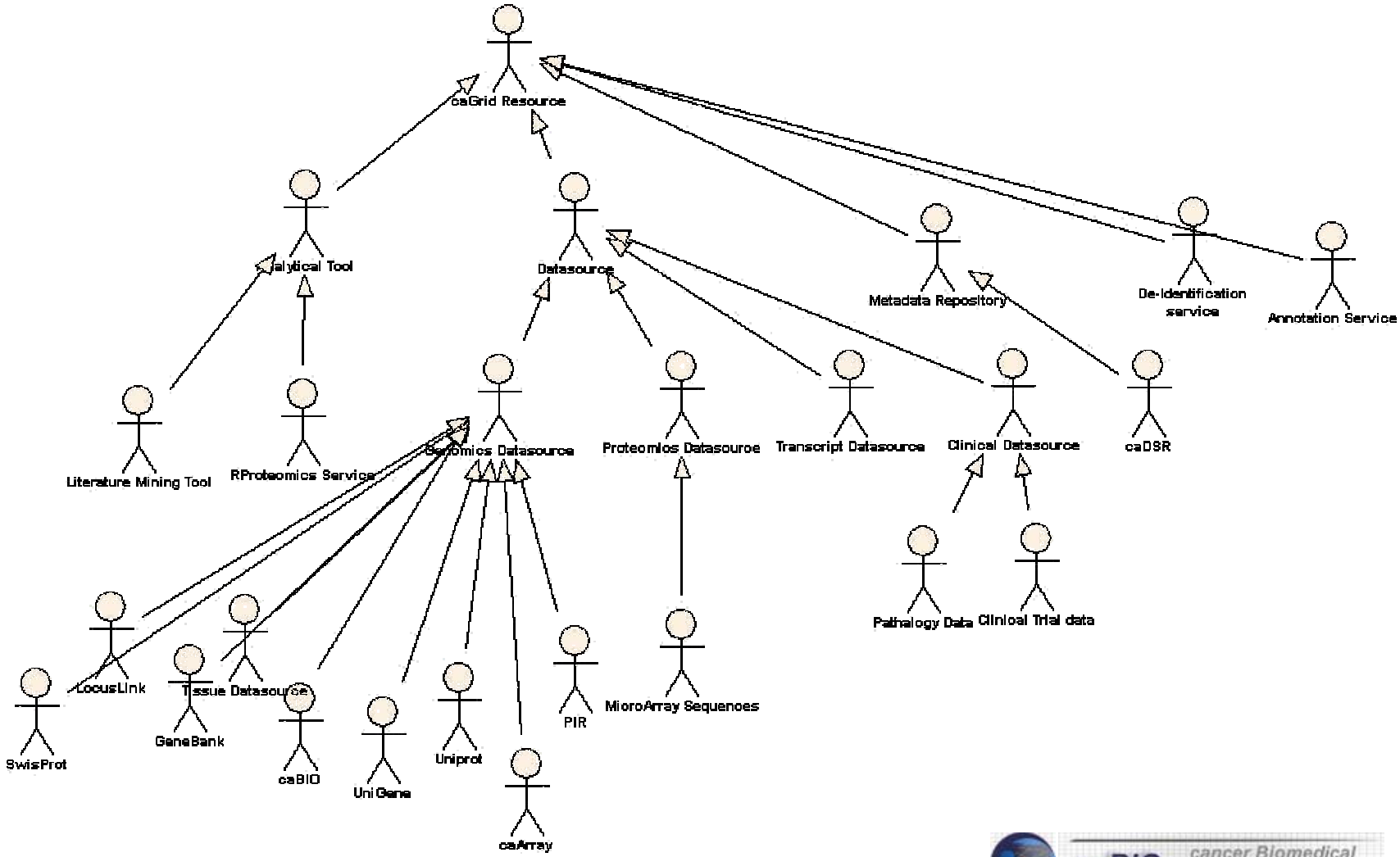
Some comments from ICR group

- Security
  - Important but separate issue.
  - Important for some data providers, but not others
- Workflow
  - Not a priority.
- Performance
  - The faster the better but not an issue.
  - Fast ID resolution was brought up
  - Performance Criteria has not yet been established

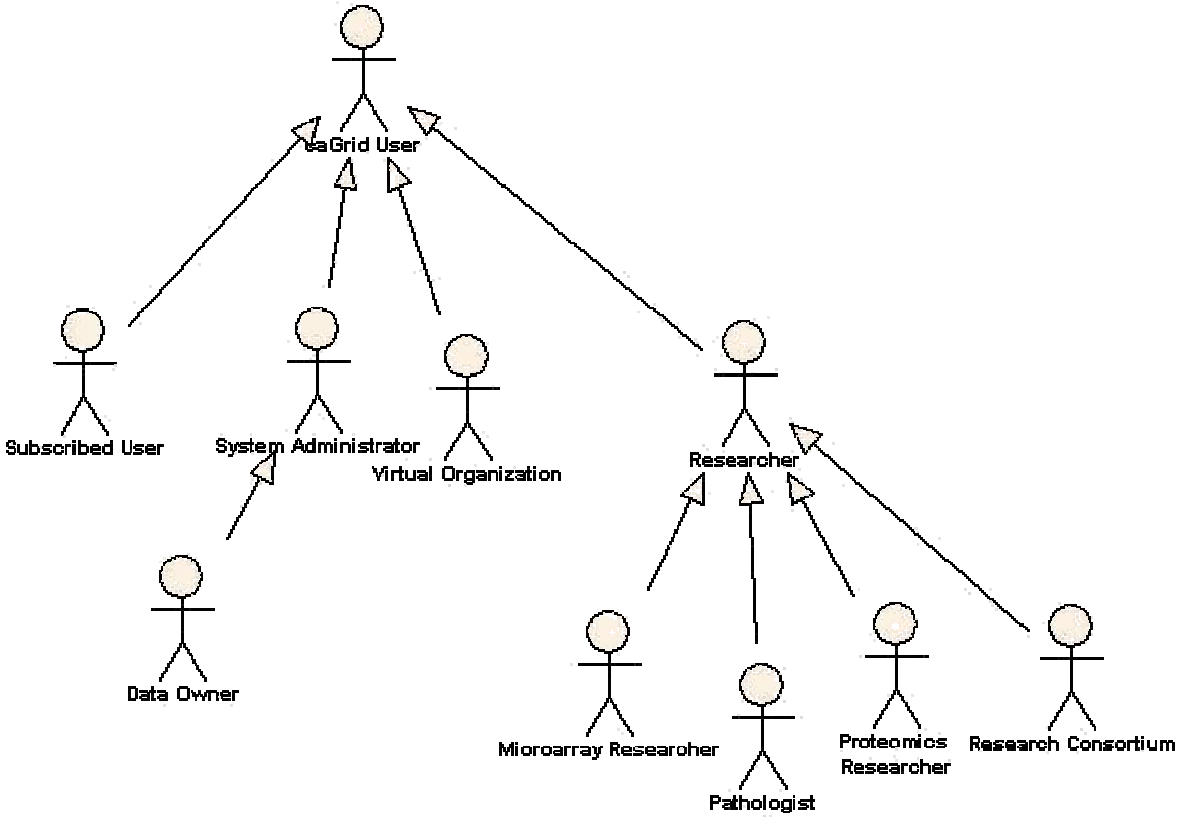
Some comments from ICR group

- Provenance
  - Difficult because you need to define the context.
  - Probably not a attribute level.
- Prototype
  - It should include Query, Analysis and Annotation.
  - First data retrieval, then ID resolution and then track version.
- Dynamic data source schema.
- Integrate public domain data source in the grid.

caGrid Actors - Resources



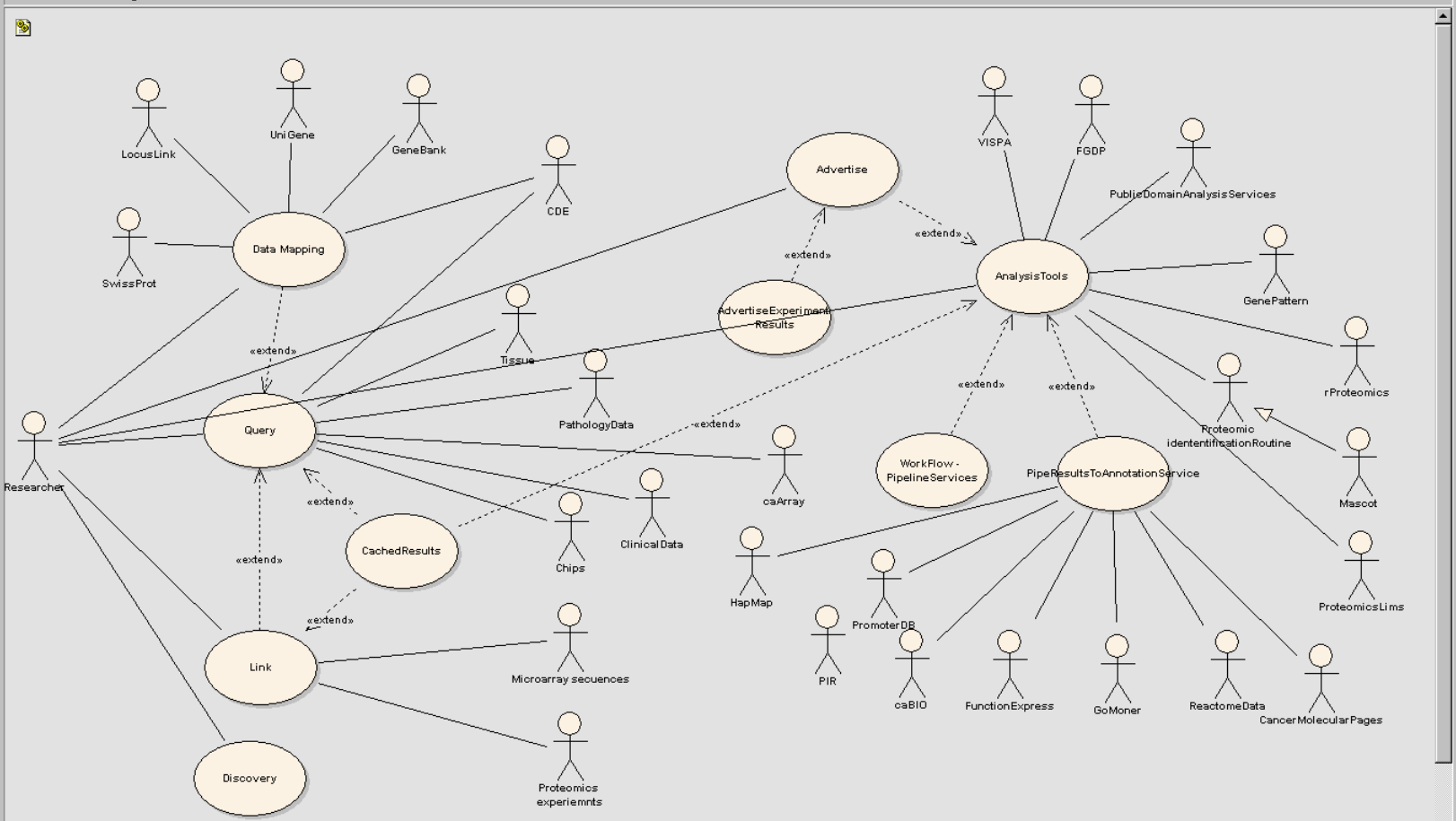
caGrid Actors - Users



[caGRID]

File Edit View Project Diagram Element Tools Configuration Help

Use Case Diagram: "Iteration 2" created: 10/18/2004 modified: 10/23/2004 10:36:03 AM 100% 799 x 1024



Start Page | Washington | Iteration 2

Project View

- Use Case Model
  - Common
  - OSU
  - Panther
  - SAIC
    - Actors
    - City of Hope (CT
    - Georgetown
    - UPENN
    - Wake Forest Un
    - Washington
      - Iteration 2
      - Washington
      - Integration
        - Actor1
        - CDE
        - CancerMolec
        - Chips
        - ClinicalData
        - FGDP
        - FunctionExpr
        - GeneBank
        - GenePatterr
        - GoMoner
        - HapMap
        - LocusLink
        - Mascot
        - Microarray s
        - PIR
        - PathologyDe
        - PromoterDB
        - Proteomic id
        - Proteomics e
        - ProteomicsL
        - PublicDomai
        - ReactomeDa
        - Researcher
        - SwissProt
        - Tissue
        - UniGene
        - VISPA
        - caArray
        - caBIO
        - rProteomics
        - Advertise
        - AdvertiseExp
        - AnalysisTool
        - CachedResu
        - CancerMolec

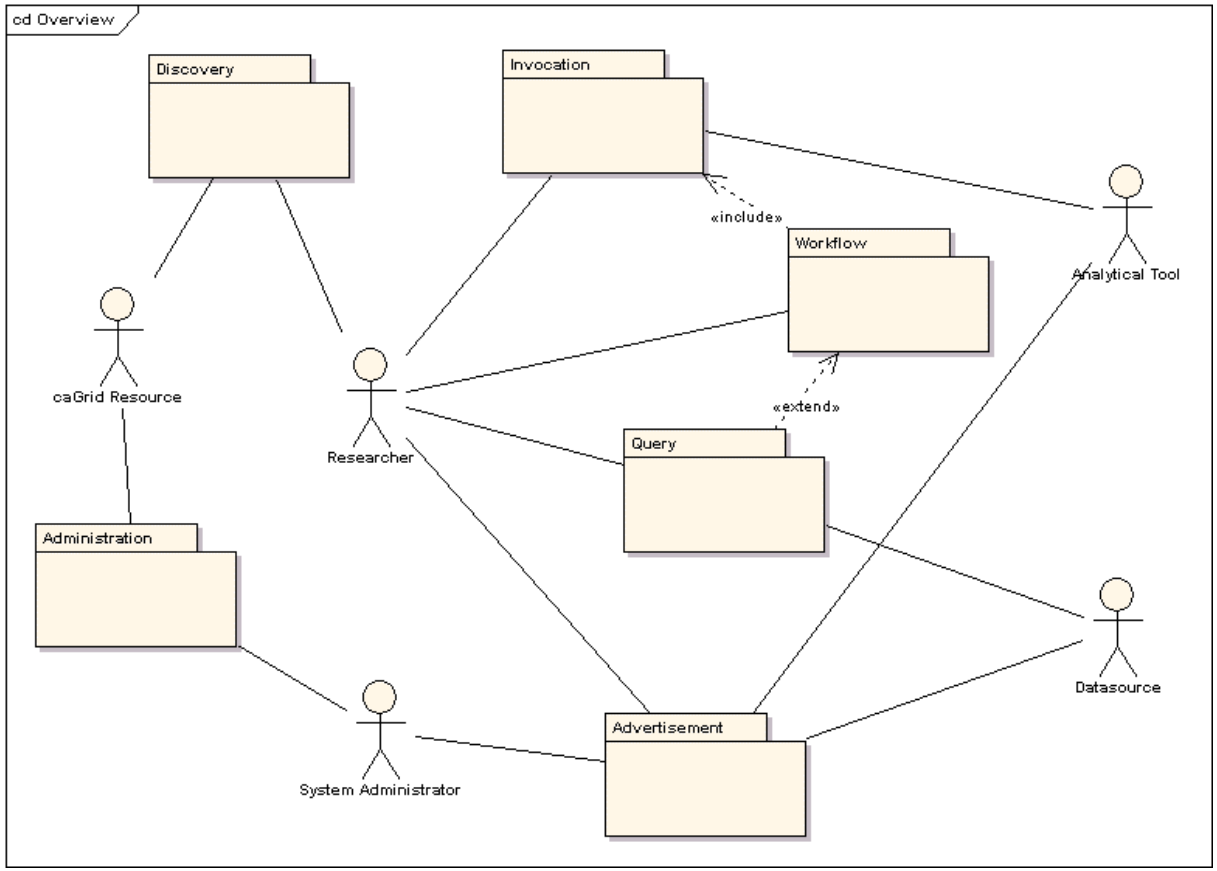
Pro... Notes Proj... Res...

System

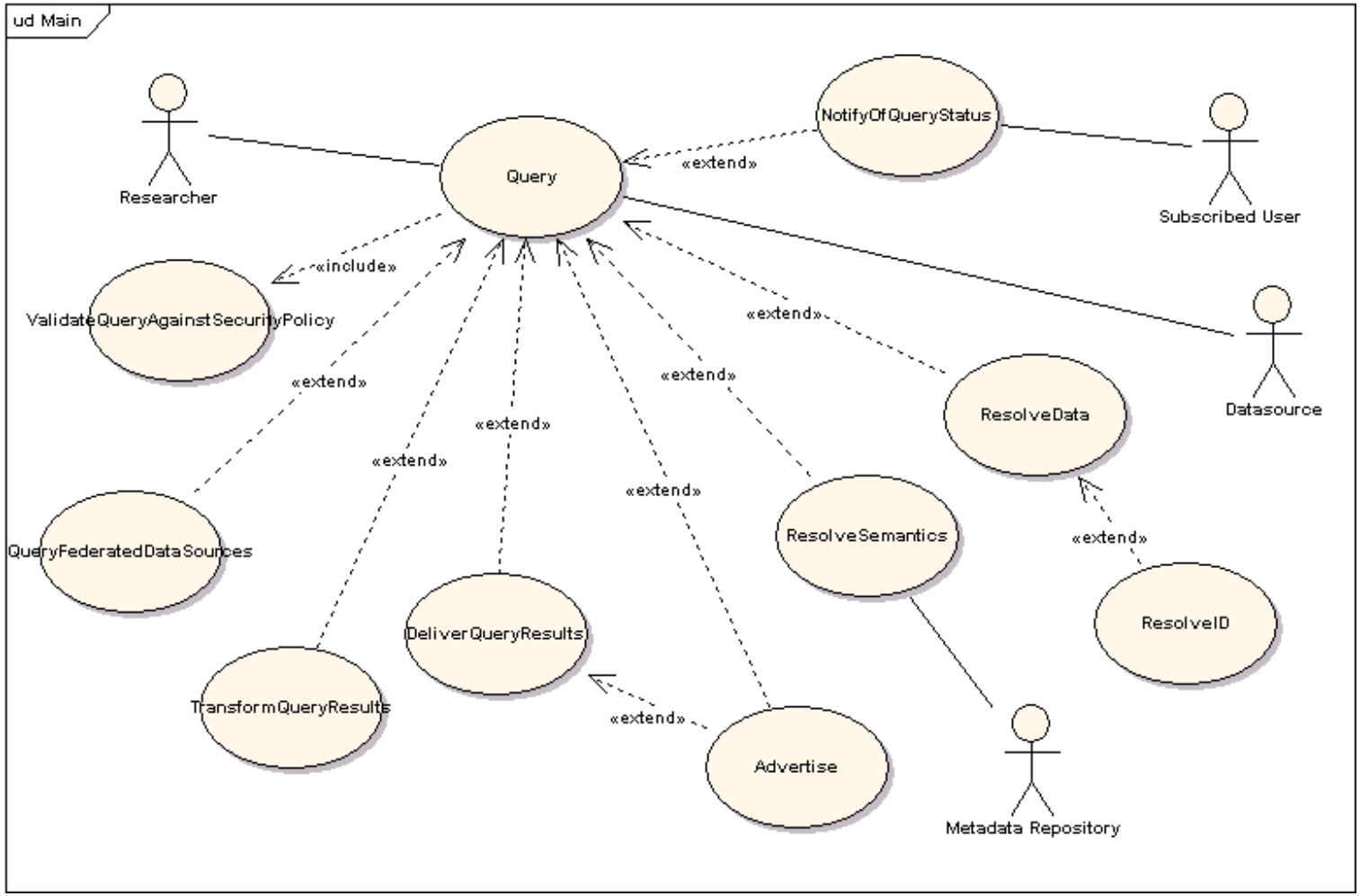
Save the current diagram

CAP NUM SCRL

Integrated Use case Model



Integrated Use case Model (Cont.)



caBIG integrated non-functional requirements.

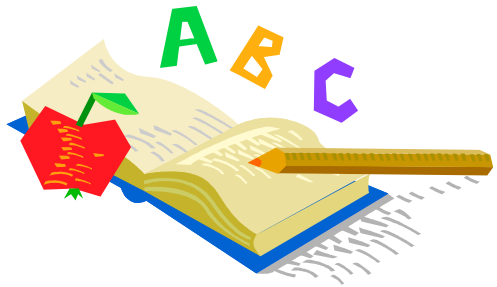
	ICR	TBP	CT	V-CDE
Security	L	H	H	N/A
Scalability	M	M	M	M
Resource management	H	H	L	L
Sys. properties	M	M	M	H
Group support	H	H	H	H
Monitoring	H	H	M	L
Connectivity	H	H	H	H
Decentralization	H	H	H	M
Data integrity	H	H	H	H
Quality of services	TBD	TBD	TBD	TBD
Industry / NCI-caBIG Standards	H	H	H	H

L: Low – M: Medium – H: High

2. Architecturally significant technologies

- **W3C - Web services, Semantic Web.**
- **GGF/OASIS - OGSA/OGSI/WS-RF.**
- **Semantic Grid.**

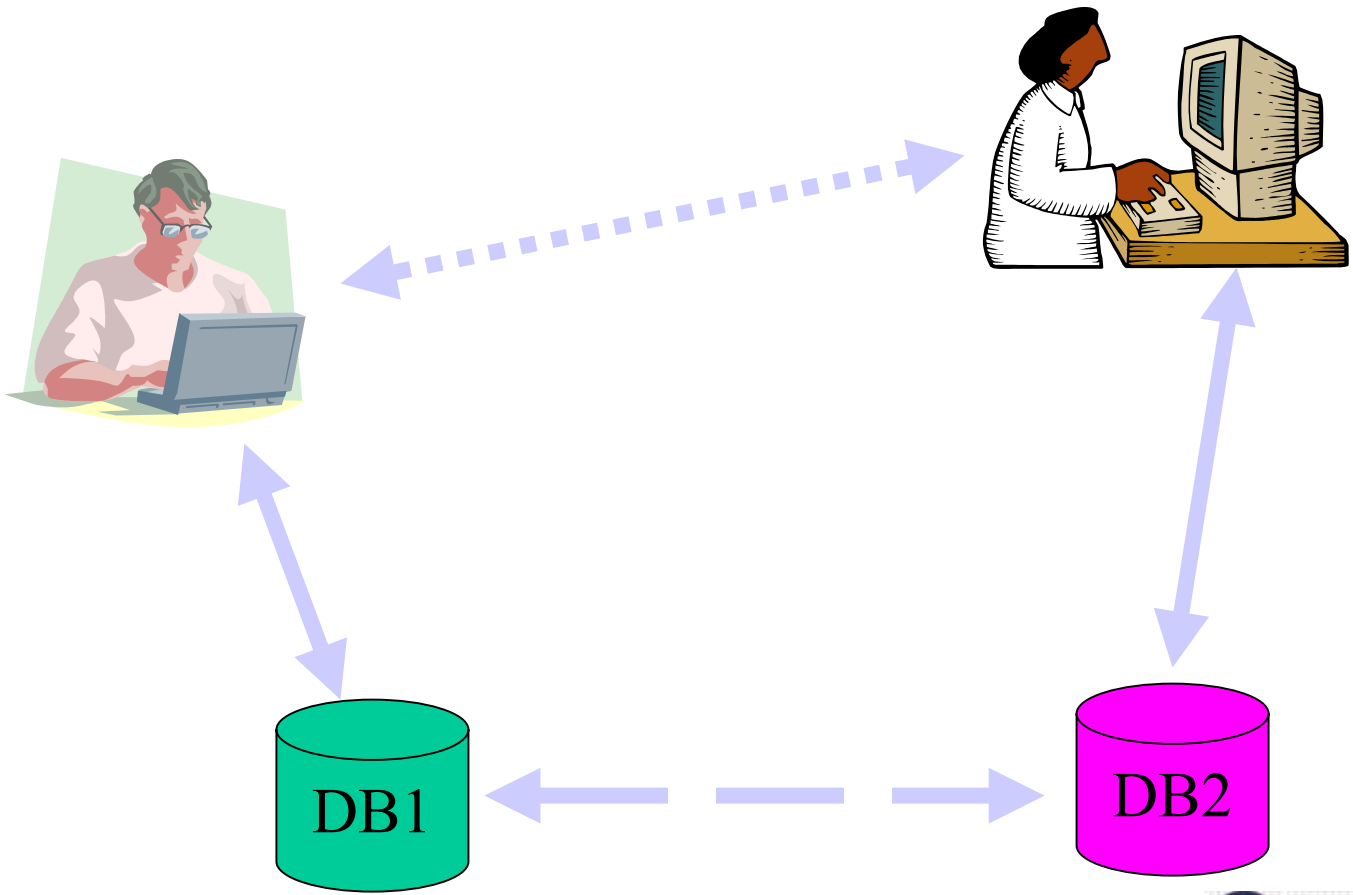
W3C - Web services, Semantic Web



Can Information be viewed as a Knowledge Asset?

Sharing Knowledge

- What needs to be exchanged and how should this be captured?



How do Scientists Assemble Knowledge

- **Scientific Annotations on one's Research**

  Every group has a different language...

- **Hypotheses and Models**

  Can we be more exact in what we're trying to say?

- **Scientific Literature ePublishing**

  Can scientific content be made machine-readable?

- **Querying vs. Aggregating**

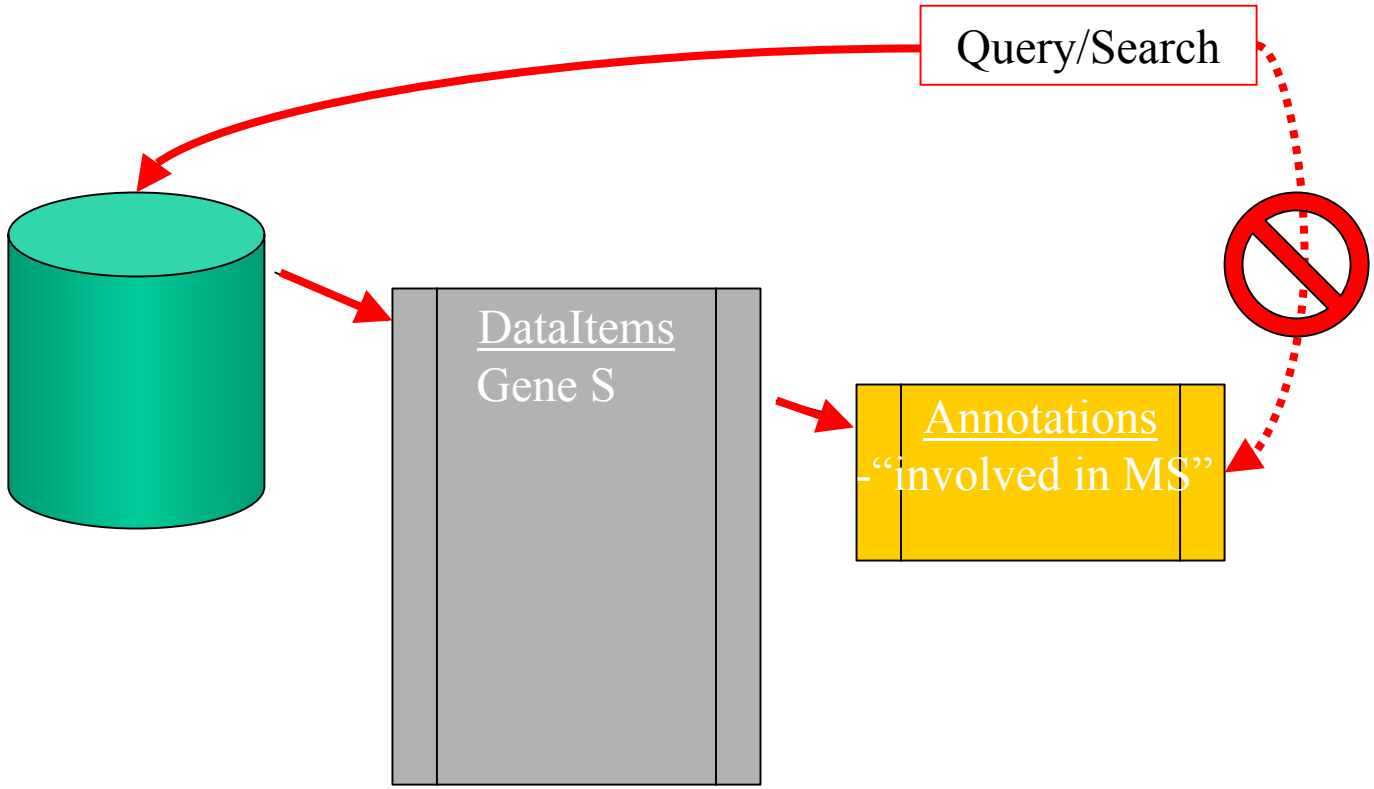
  Databases linked by ontologies and the Web

Annotations

Noting something worthwhile for others to see

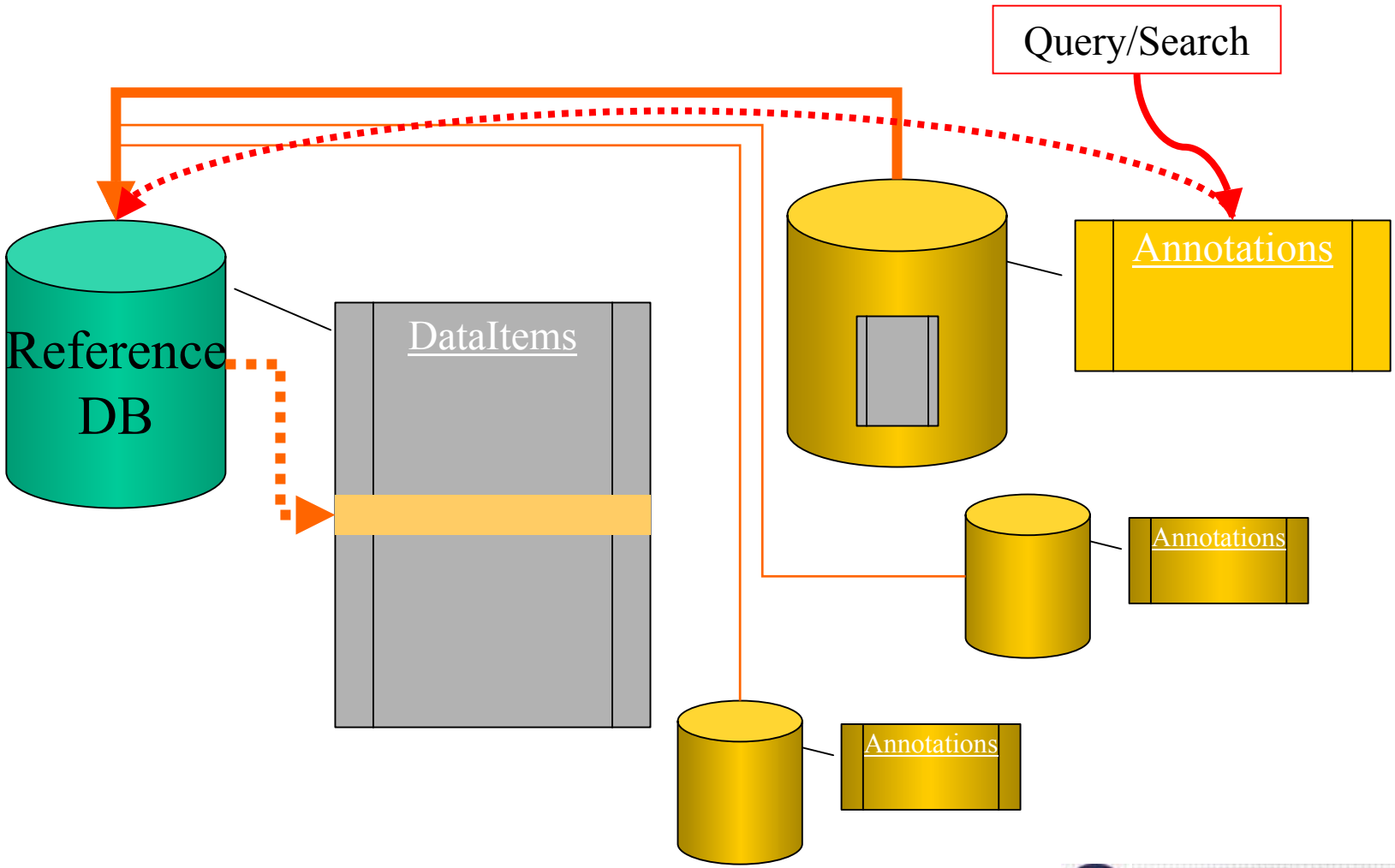
Annotations

– saying something worthwhile for others to see

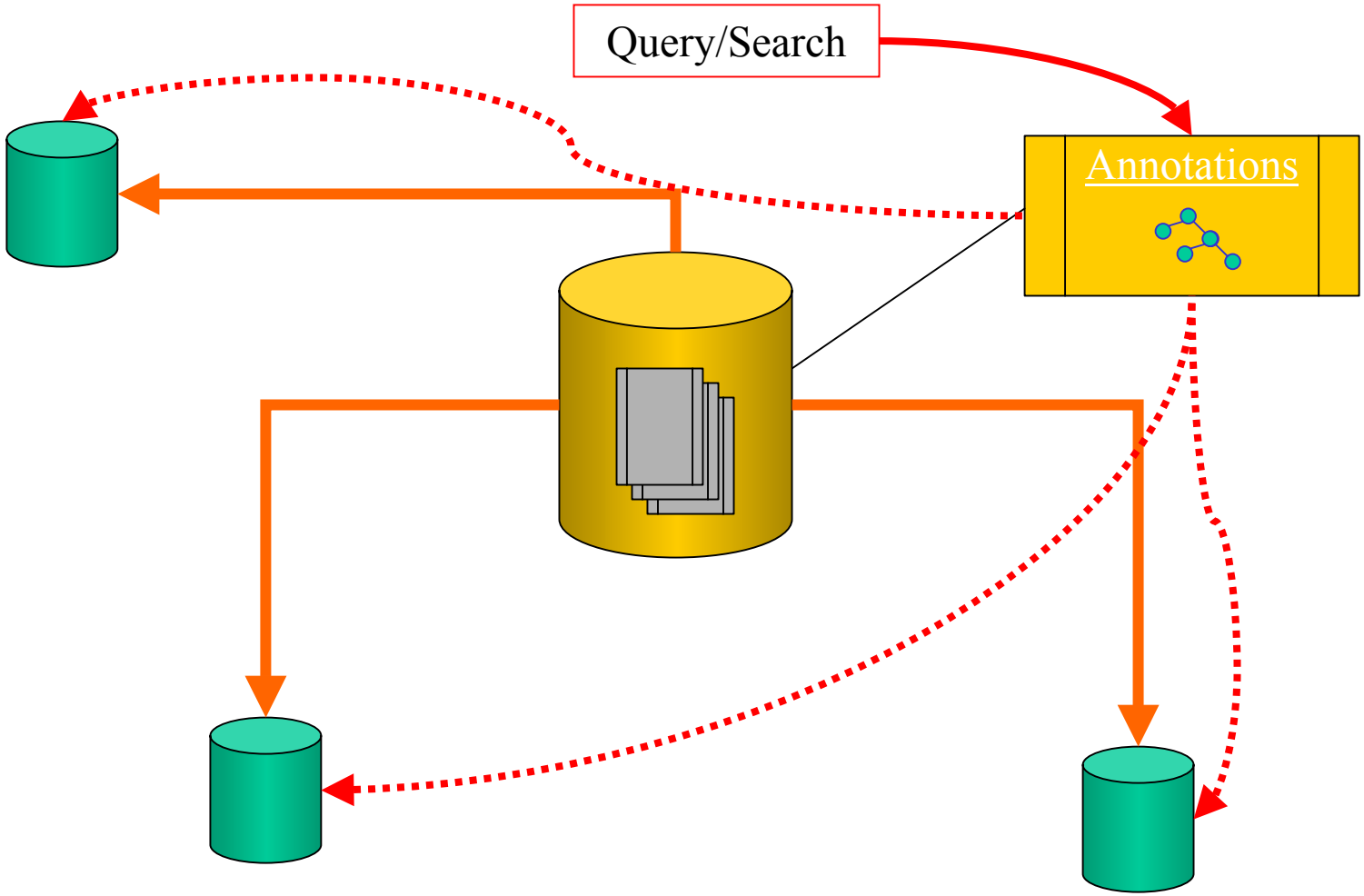


Annotations ala DAS and Web Services (Lincoln Stein/Brian Gilman)

– Distributed Annotations to a common Reference DB



From Annotations to Aggregations



Embedded Semantics

<novel_method><predicting><therapeutic_group>

Prediction of clinical drug efficacy by classification of drug-induced genomic expression profiles *in vitro*

Erik C. Gunther*†, David J. Stone*, Robert W. Gerwien, Patricia Bento, and Melvyn P. Heyes

CuraGen Corporation, 322 East Main Street, Branford, CT 06405

Edited by Floyd E. Bloom, The Scripps Research Institute, La Jolla, CA, and approved June 11, 2003 (received for review April 30, 2003)

Assays of drug action typically evaluate biochemical activity. However, accurately matching therapeutic efficacy with biochemical activity is a challenge. High-content cellular assays seek to bridge this gap by capturing broad information about the cellular physiology of drug action. Here, we present a method of predicting the general therapeutic classes into which various psychoactive drugs fall, based on high-content statistical categorization of gene expression profiles induced by these drugs. When we used the classification tree and random forest supervised classification algorithms to analyze microarray data, we derived general "efficacy profiles" of biomarker gene expression that correlate with antidepressant, antipsychotic and opioid drug action on primary human neurons *in vitro*. These profiles were used as predictive models to classify naïve *in vitro* drug treatments with 83.3% (random forest) and 88.9% (classification tree) accuracy. Thus, the detailed information contained in genomic expression data is sufficient to match the physiological effect of a novel drug at the cellular level with its clinical relevance. This capacity to identify therapeutic efficacy on the basis of gene expression signatures *in vitro* has potential utility in drug discovery and drug target validation.

tion tree (CT) and random forest (RF) supervised classification schemes can be used to predict the functional category of members of each of these drug classes with good accuracy, based on analysis of the gene expression profile induced by a drug.

Materials and Methods

Cell Cultures. Primary human neuronal precursor cells (Clonexpress, Gaithersburg, MD) were cultured for 7 days in growth media (GM) (50:50 DMEM/F12, 5% FBS, 10 ng/ml basic fibroblast growth factor, 10 ng/ml epidermal growth factor, 1:100 Clonexpress neuronal cell supplement, penicillin/streptomycin) and differentiated for 7 days in six-well plates at 900,000 cells per well in GM plus 100 μM dibutyrtyl cAMP, 20 ng/ml nerve growth factor, 1:100 matrigel, with 72-h media changes. Morphologically neuronal cells comprised ≈80% of the cultures.

Drug Treatments. Drugs were dissolved in DMSO and added to cultures at a final DMSO concentration of 0.04%. Drug concentrations represented pharmaceutically relevent doses: 2.0 μM amoxepine, 2.0 μM clomipramine, 2.0 μM desipramine, 1.0

Ontologies

Semantics & Ontologies

"I've got an idea to share…

but how to I express it if we don't have the same *language* ?"

- Semantics-

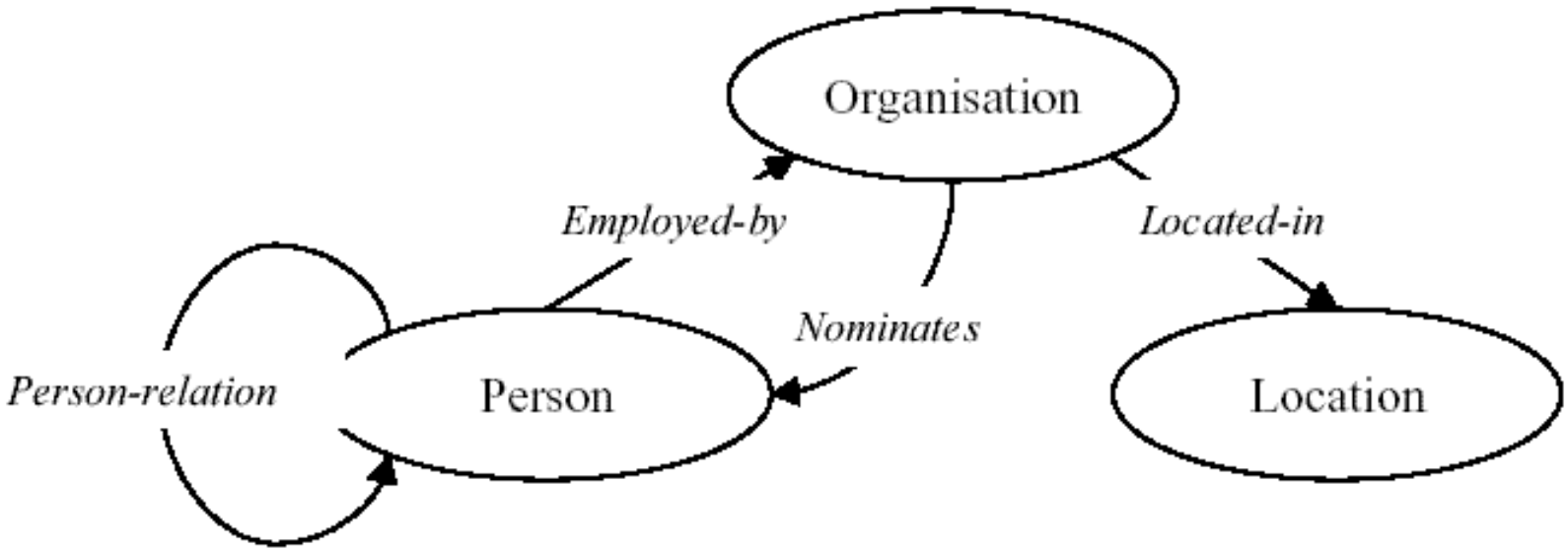
  Colorless green ideas sleep furiously

- Ontologies- Concepts, Relations, and Instances

  I just bought a mustang, but it's not running on all four!

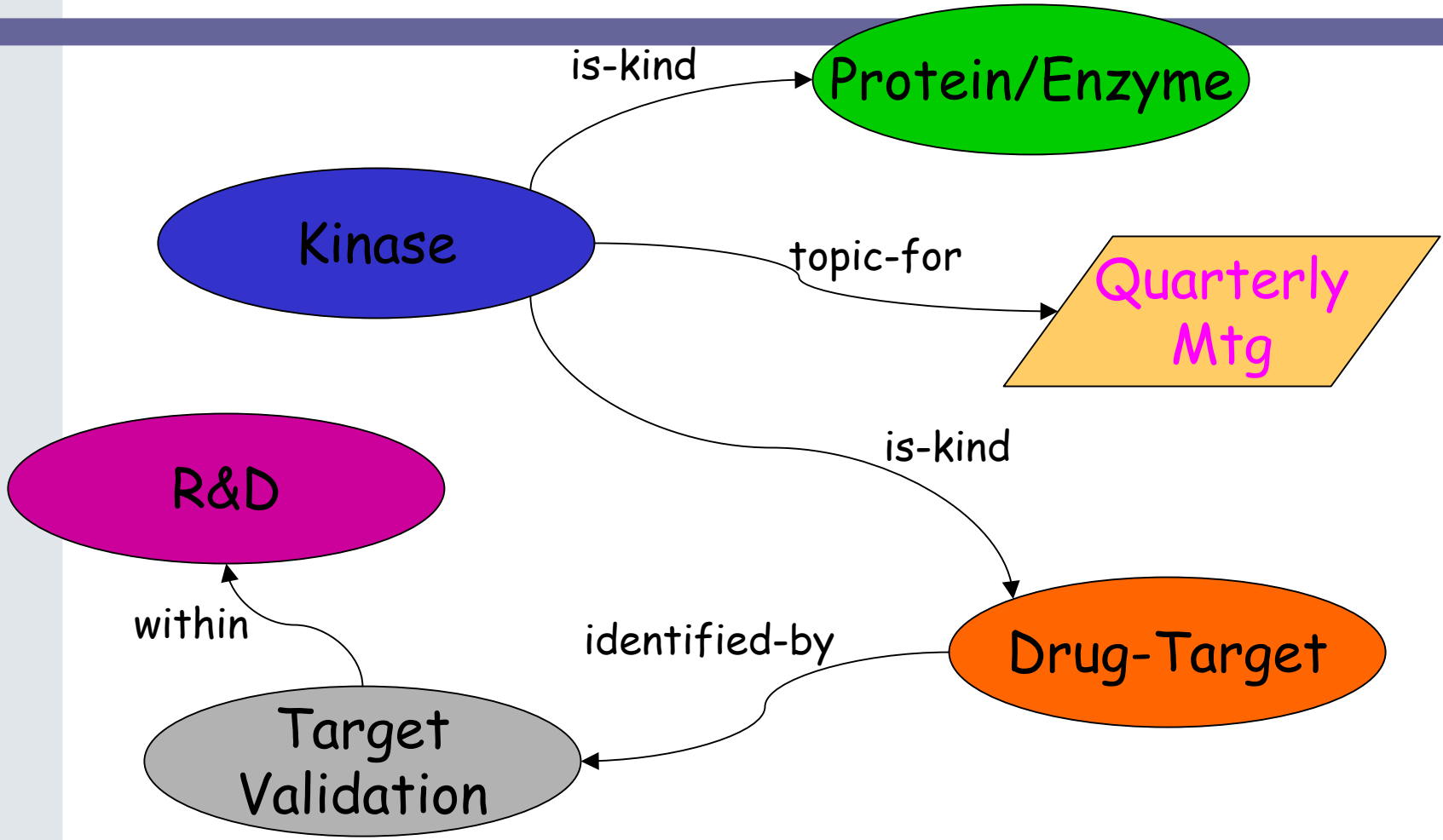
- These beyond an IT approach!

Ontologies are not ERDs



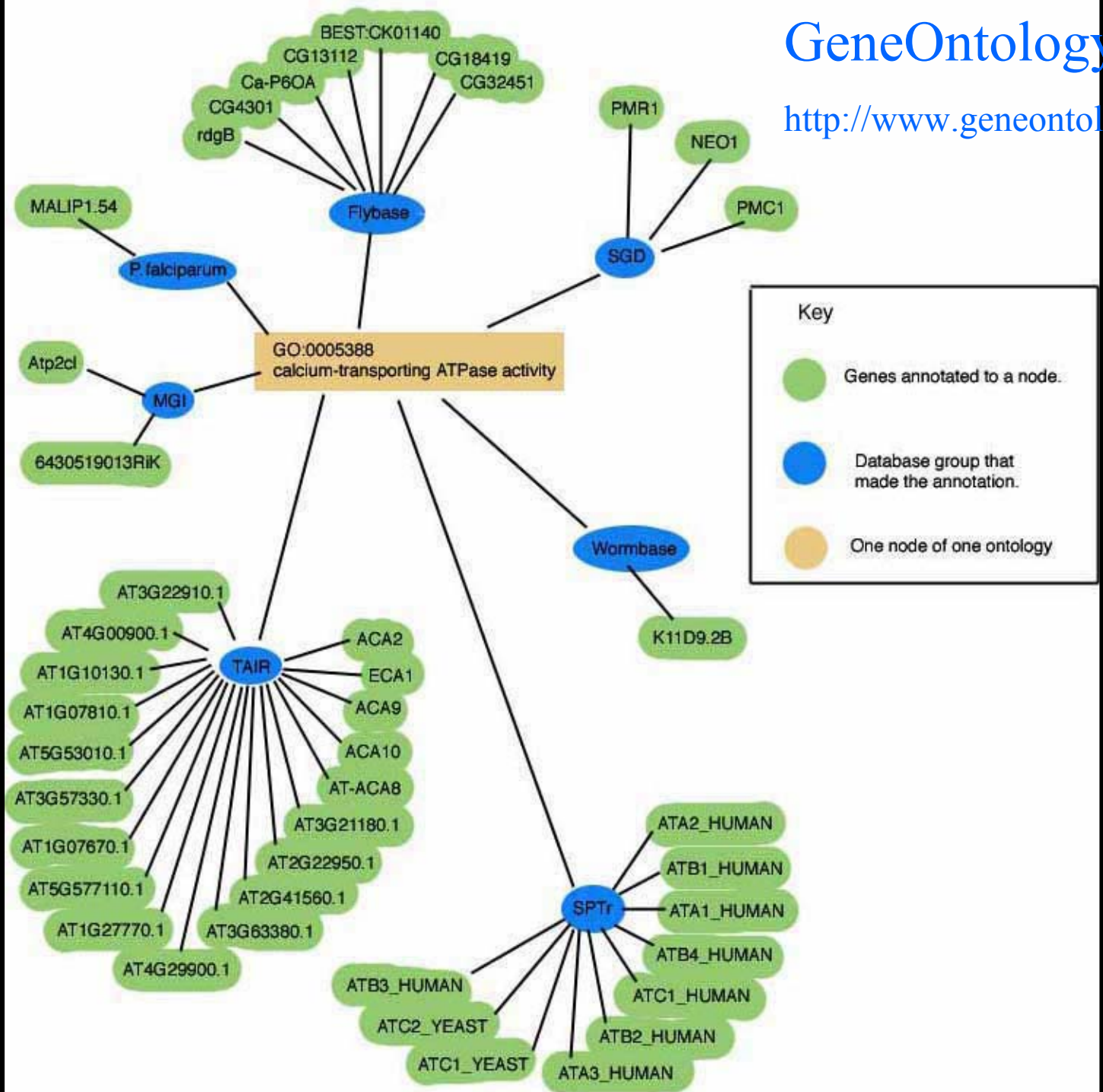
How do we go about defining them? User-driven

Ontologies –
Combining domain and business logic



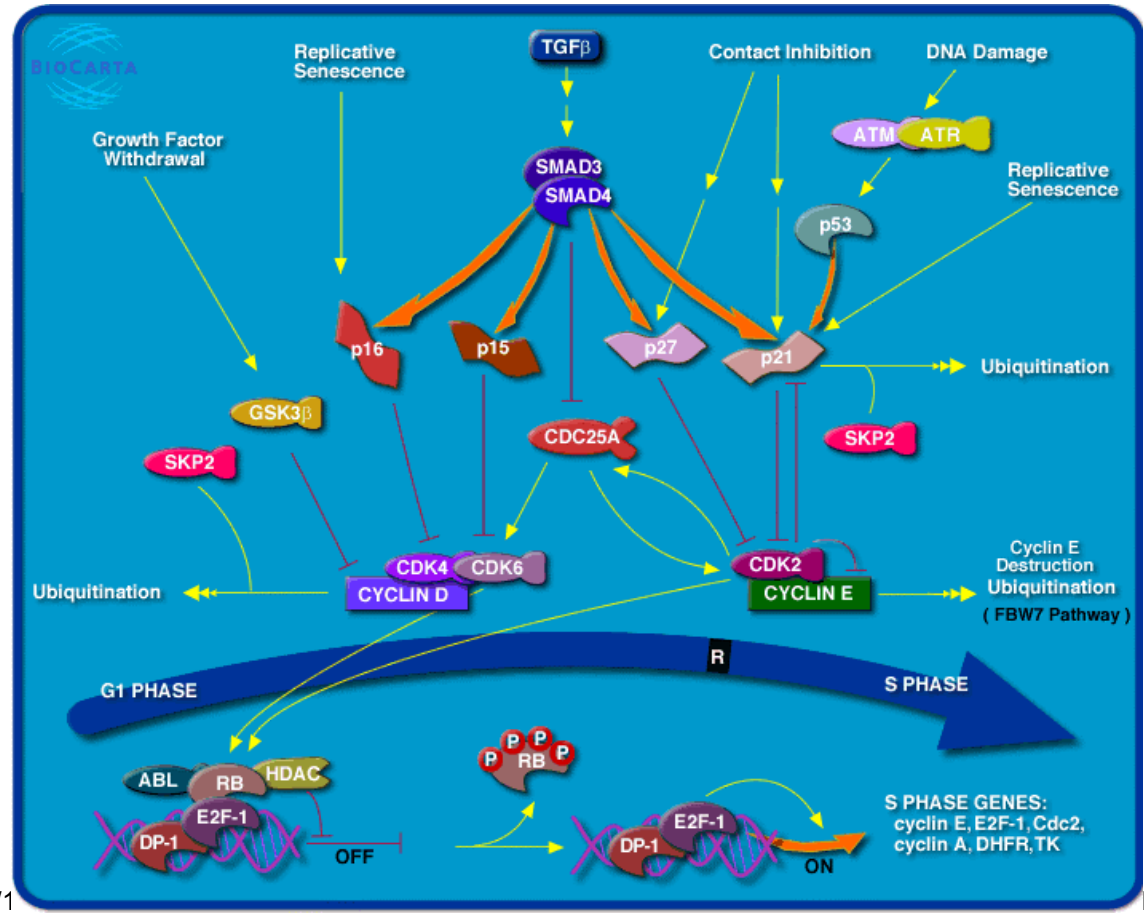
GeneOntology

http://www.geneontology.org/

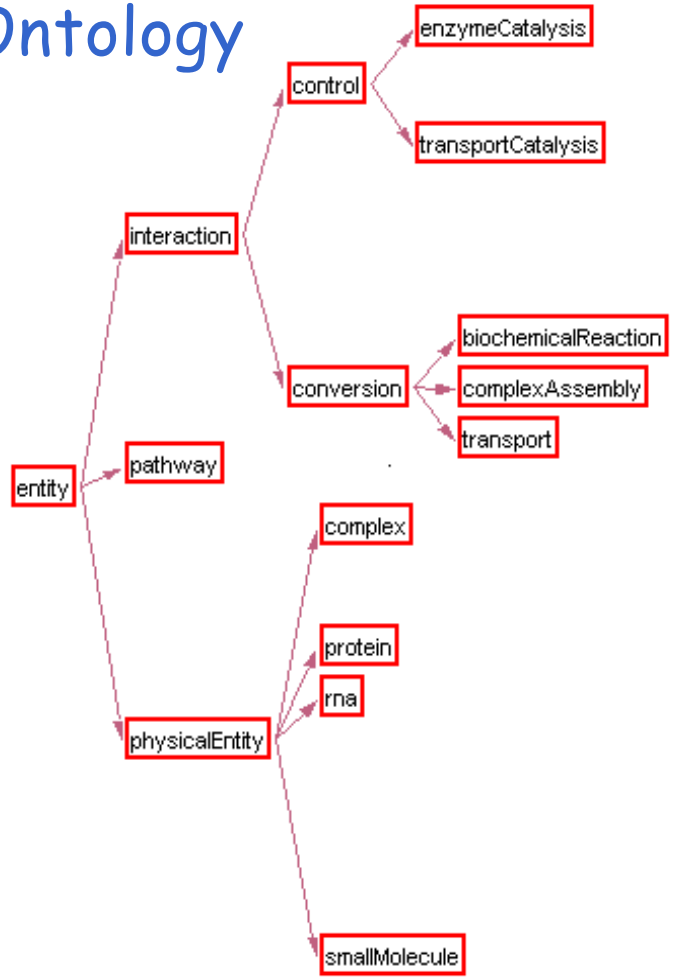


Molecular Pathways-

Models that help us find new drug targets for cancer research



BioPAX Ontology



Applications for Ontologies

- Define, associate, and share common languages between groups
- Capture all research annotations
- Dynamically aggregate disparate information and knowledge towards hypothesis building
- Newsfeeds for JIT aggregation: Nature's Urchin
- Formalize Organizational roles and processes

How Can the Semantic Web Help?

POLICY FORUM

COMMUNICATION

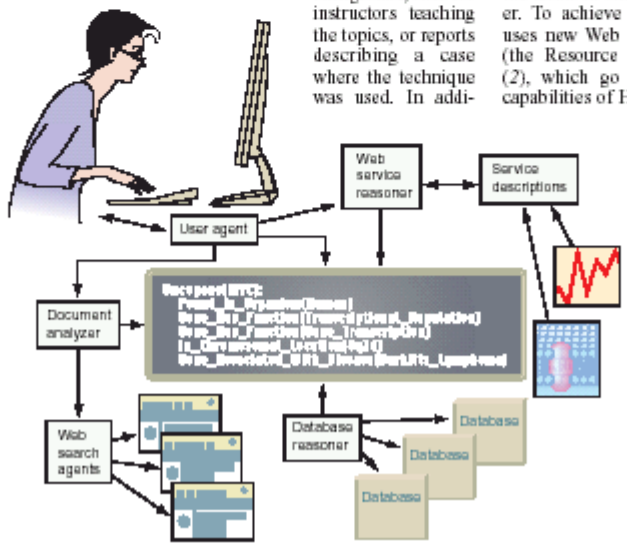
Science and the Semantic Web

James Hendler

Scientists have become increasingly reliant on the World Wide Web for supporting their research endeavors. The Web is used for finding preprints and papers in online repositories, for participating in online discussions at sites such as *Science Online*, for accessing databases through specialized Web interfaces, and even for ordering scientific supplies. When searching for a specific Web site or a paper on a particular topic, engines like Google can do a phenomenal job of sorting through billions of possibilities and identifying potentially useful candidates, often within the first few search results. On specialized Web sites, domain-specific search engines can do even better, for example, enabling the mathematician to easily find papers on "symplectic geometry" or the physicist to see preprints relating to "mesoscopic systems and the quantum hall effect." In fact, the Web has become indispensible for supporting the traditional communications within our disciplines and the needs of scientists within their disciplinary boundaries.

Enhanced online at www.sciencemag.org/cgi/content/full/299/5606/520

However, as modern science continues its exponential growth in complexity and ferent scientific field. On the Web, however, scientists looking for results in sites developed for different scientific communities are often at a loss. For example, a scientist searching for a technique to analyze some image-based data may not know to look for papers on Laplacean invariants (found under the symplectic geometry category in many math sites). A general search on image analysis will find thousands of possibilities but will provide little or no guidance as to which sites can explain how to use the techniques, as opposed to finding papers formalizing the mathematical background, sites for instructors teaching the topics, or reports describing a case where the technique was used. In addition, the Web is even more limited when



information technologists must forge new models of cooperation, and new thinking must go into the funding and dissemination of this next generation of scientific tools on the Web.

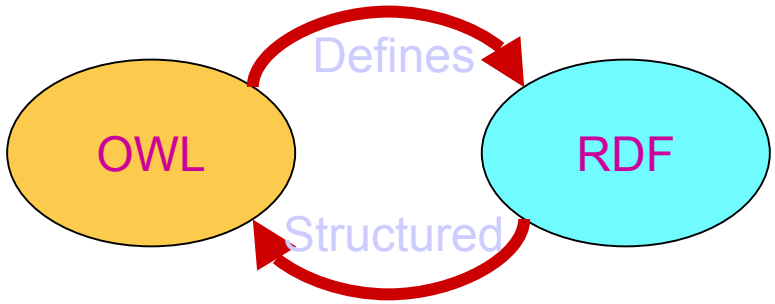
A new generation of Web technology, called the Semantic Web (*1*), is designed to improve communications between people using differing terminologies, to extend the interoperability of databases, to provide tools for interacting with multimedia collections, and to provide new mechanisms for the support of "agent-based" computing in which people and machines work more interactively.

Whereas the current Web provides links between pages that are designed for human consumption, the Semantic Web augments this with pages designed to contain machine-readable descriptions of Web pages and other Web resources. These documents can be linked together to provide information to the computer as to how the terms in one relate to those in another. To achieve this, the Semantic Web uses new Web languages based on RDF (the Resource Description Framework) (*2*), which go beyond the presentation capabilities of HTML (Hypertext Markup Language, which is used on most Web sites today) and the document-tagging capabilities of the Extensible Markup Language (XML), a more recent innovation being used to allow parts of documents to be more precisely delineated.

The Center for Bioinformatics of the U.S. National Cancer Institute (NCI), as part of the Metathesaurus project (*3*), is turning a large vocabulary of cancer research terms into a machine-readable "ontology"—essentially an expanded thesaurus that delineates precise relationships between the vocabulary items

Framework for Next Generation of the Web
Knowledge Exchange within a Semantic Web

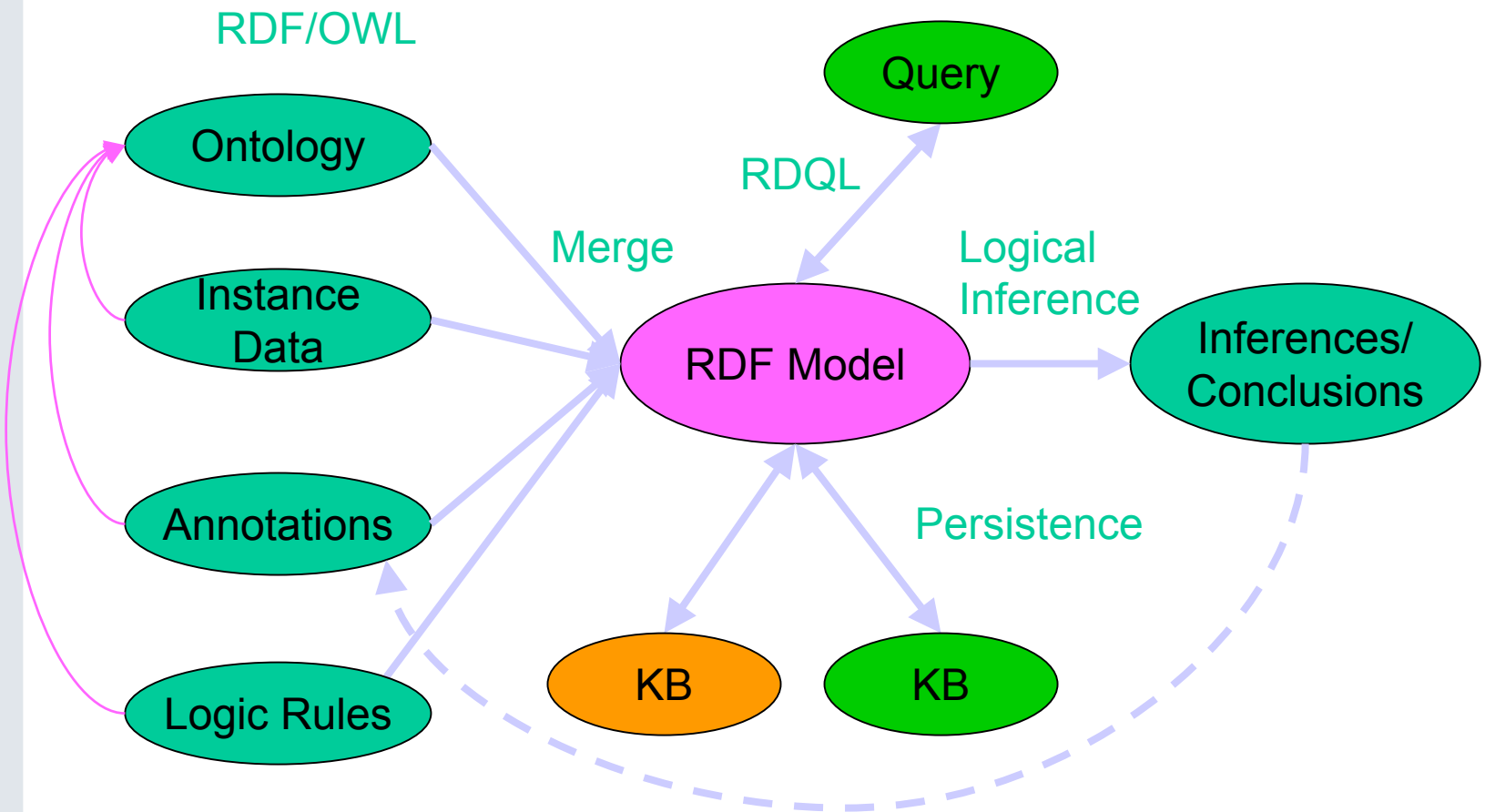
- OWL (Ontology Web Language)
  - W3C Ontology Specification
  - Goes beyond 1st order Logic (Frames & Descriptive Logic)
  - Extensible by members of any community
  - Structurally based on RDF
- RDF (Resource Description Framework)
  - Basic XML Semantic Format that OWL is based upon
  - Allows users to merge and aggregate any set of related data and relational components
  - Refers to Ontologies specified in OWL



OWL-RDF – *more expressive than XML-Schema*

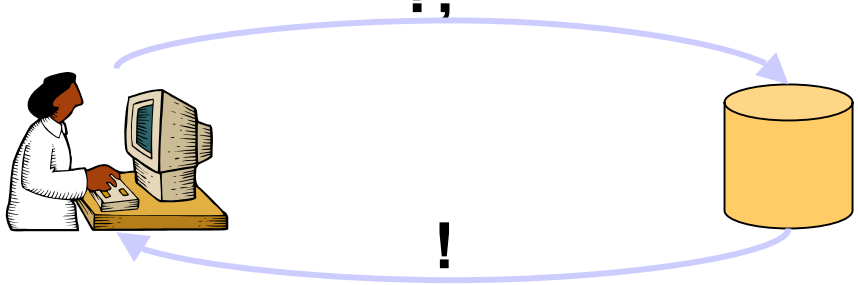
```
<owl:Class rdf:about="http://www.bg.org/bio#BioChemical">
    <owl:hasProperty rdf:resource="http://www.bg.org/bio#binds"/>
    <owl:hasProperty rdf:resource="http://www.bg.org/bio#isDownStream"/>
    <owl:hasProperty rdf:resource="http://www.bg.org/bio#isUpStream"/>
</owl:Class>

<owl:Class rdf:about="http://www.bg.org/bio#BioComplex">
    <owl:hasProperty rdf:resource="http://www.bg.org/bio#composedOf"/>
    <owl:subClassOf rdf:resource="http://www.bg.org/bio#BioChemical"/>
</owl:Class>

<owl:Class rdf:about="http://www.bg.org/bio#Enzyme">
    <owl:hasProperty rdf:resource="http://www.bg.org/bio#catalyzes"/>
    <owl:subClassOf rdf:resource="http://www.bg.org/bio#Protein"/>

<owl:Class rdf:about="http://www.bg.org/bio#Compound">
    <owl:subClassOf rdf:resource="http://www.bg.org/bio#BioChemical"/>
</owl:Class>
```

An RDF Aggregation Use-Case

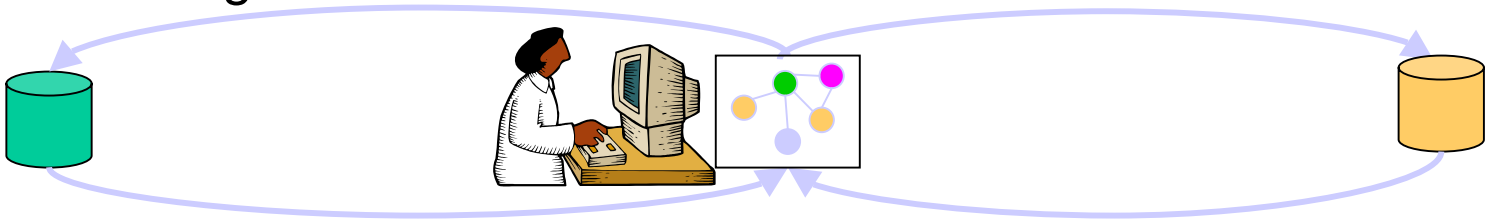


New Data Paradigm for Research

- More than a collection of tables for Set-selection

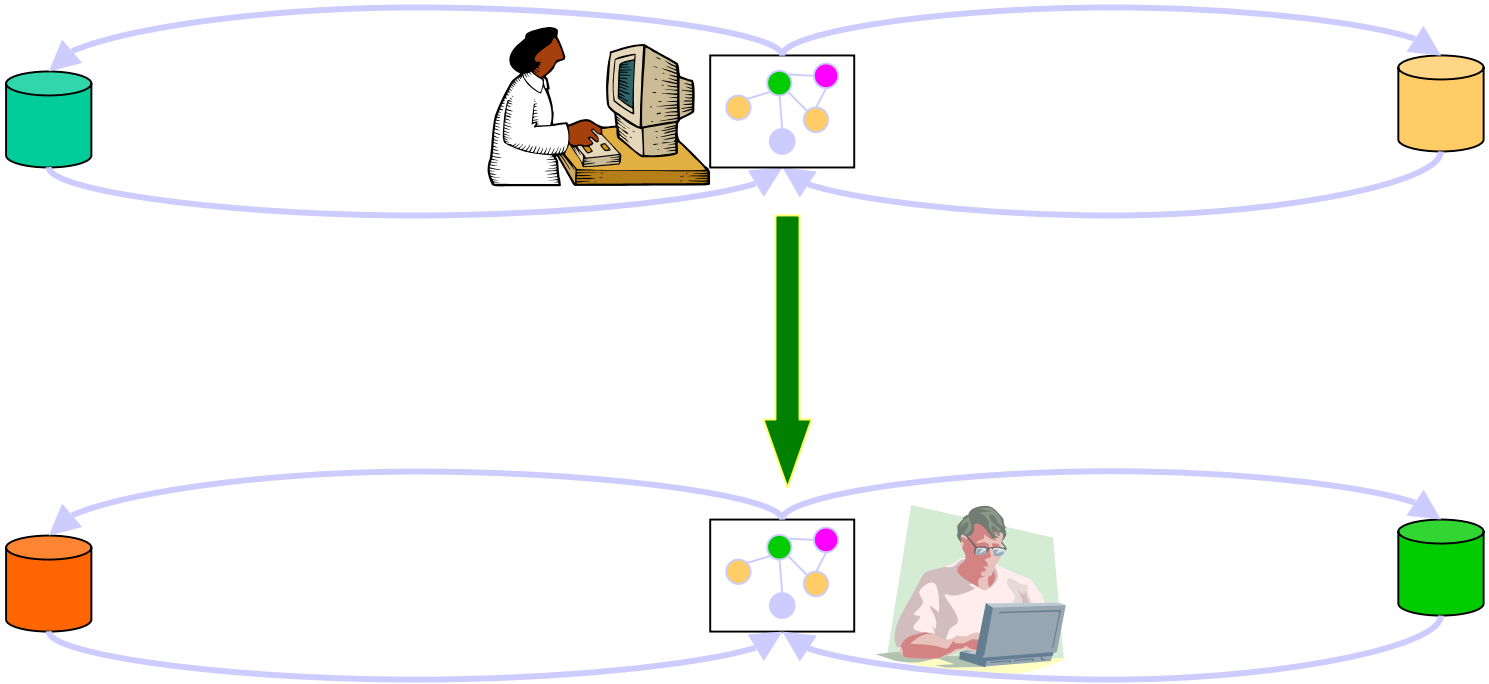


- Data can evolve with additions of attributes and properties as well as through new inferences

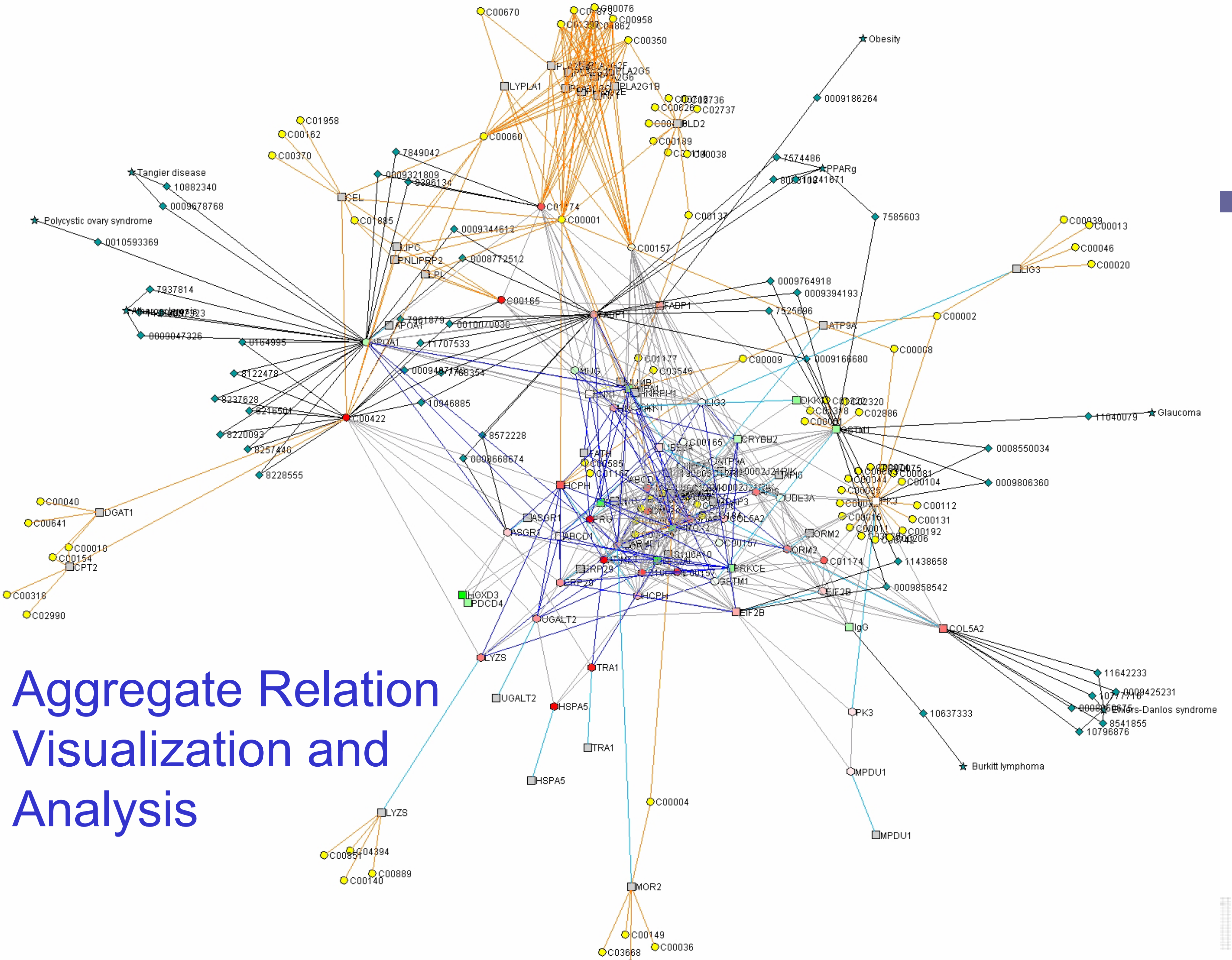


New Data Paradigm for Research

- Sharing discoveries in context



Aggregate Relation Visualization and Analysis



RDF- Aggregation Example

```
@prefix nlm: <http://www.nlm.nih.gov/diseases-
schema#> .bg:OncoGene a owl :Class ; owl:subClassOf
bg:Gene .

bg:FusedOncoGene a owl :Class ; owl:subClassOf
bg:OncoGene .

:gAbl_TK a bg:OncoGene ;
          bg:hasProduct :pAbl_TK ;
          bg:hasTranscript :mAbl_TK ;
          nlm:lsid "gi238783" .

:gBCR_Abl_TK a bg:FusedOncoGene ;
          bg:hasProduct :pBCR_Abl_TK ;
          bg:composedOf :gAbl_TK ; bg:composedOf :Bcr
;
          bg:expressedIn nlm:Myeloid ;
          bg:isImplicatedIn :CML ;
          bg:Comment "Chimeric Gene" ;
          nlm:lsid "gi282887" .

:Gleevec a bg:Drug ;
          bg:id "STI571" ;
```

Facts: Abl is a kinase, that is involved in CML when it is rearranged with BCR to form a fused-gene. Gleevec targets this protein.

Infer: Gleevec may reverse CML.

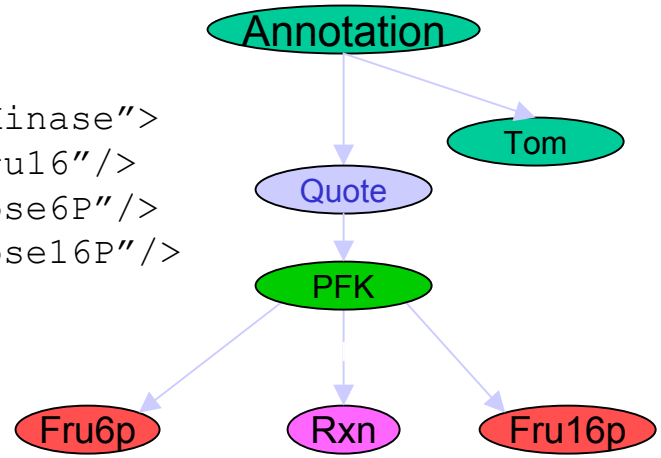
RDF- Inferencing from Distributed Facts

```
:Abl      bg:biomarkersFor :CML .

:Bcr      bg:isImplicatedIn :CML .

:CML      bg:D_perturbs :STKinaseCascade .

:Gleevec      bg:affectsTissue nlm:Myeloid;
      bg:mayCure :CML;
      bg:targets :gAbl_TK, :gBCR_Abl_TK .

:JNK      bg:biomarkersFor :CML .

:Mcl-1      bg:biomarkersFor :CML .

:STAT5      bg:biomarkersFor :CML .

:cyclin-D1      bg:biomarkersFor :CML .

:gAbl_TK      a bg:Gene;
      bg:alt_target :CML;
      bg:isImplicatedIn :CML .

:gBCR_Abl_TK      a bg:Gene,
            bg:OncoGene .
```

Fine-Grain Annotations—

"Tom specifies that PFK catalyzes the reaction of Fru6p to Fry1,6p"

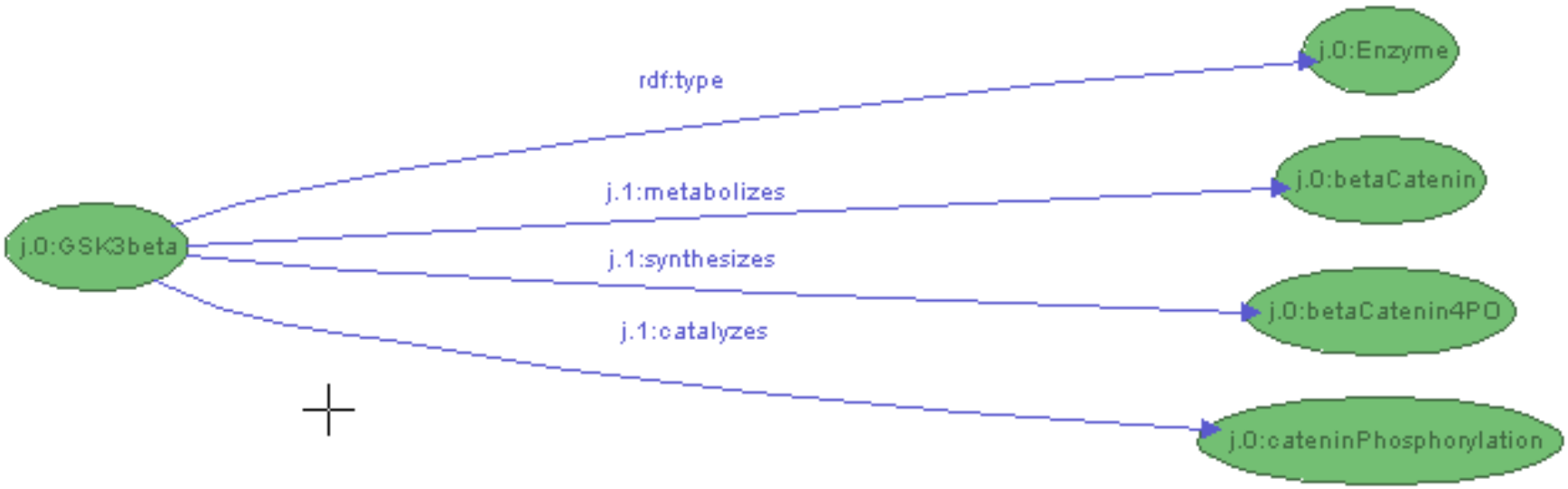
```
<rdf:RDF xmlns="file:/Python22/CWM/anno.rdf#"
    xmlns:dc="http://purl.org/dc/elements/1.1/"
    xmlns:bg="http://www.bg.org/"
    xmlns:log="http://www.w3.org/2000/10/swap/log#"
    xmlns:owl="http://www.w3.org/2002/07/owl#"
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
    xmlns:hupo="http://www.hupo.org/proteins#"
    xmlns:cas="http://www.cas.org/chemicals#"  >


  <bg:Annotation >
     <dc:author>Tom Plasterer</dc:author>
     <bg:specifies rdf:parseType="Quote">
        <bg:Enzyme rdf:about="hupo#PhosphoFructoseKinase">
           <bg:catalyzes rdf:resource="hupo#Fru62Fru16"/>
           <bg:metabolizes rdf:resource="cas#fructose6P"/>
           <bg:synthesizes rdf:resource="cas#fructose16P"/>
         </bg:Enzyme>
     </bg:specifies>
  </bg:Annotation>
</rdf:RDF>
```



Facts and Hypotheses…

```
p:GSK3beta      a p:Enzyme;
          bio:catalyzes p:cateninPhosphorylation ;
          bio:metabolizes p:betaCatenin ;
          bio:synthesizes p:betaCatenin4PO .


:Brian bio:states {
          :Eric bio:states {
                    bio:Confidence bio:prob "0.33" .
                    p:GSK3beta   bio:kinases p:XPDG ;
                              bio:binds p:Axin .
                    p:XPDG nih:influences nih:ColonCancer .
          } .
} .
```

Facts and Hypotheses…as graph



Inference based on Hypotheses…

```
:Melissa     :believes met:Brian, :Eric, :Tom .

  :Confidence     :prob "0.33" .

  p:GSK3beta     a p:Enzyme;
        :binds p:Axin;
        :catalyzes p:cateninPhosphorylation;
        :kinases p:XPDG;
        :metabolizes p:betaCatenin;
        :synthesizes p:betaCatenin4PO .

  p:XPDG     nih:influences nih:ColonCancer .
```

Inference based on Hypotheses…

Semantic Web for Life Sciences

- New Interest Group at W3C
- Aligned with W3C's Semantic Web
  - http://www.w3.org/2001/sw/
- Not Standards, but real-world implementations and best practices
- W3C members can automatically become members

RDF Resources

- RDF Basics - http://www.w3.org/RDF
- RDF Tools
  - JENA (Java) - http://www.hpl.hp.com/semweb/
  - IsaVIZ - http://www.w3.org/2001/11/IsaViz/
  - CWM (Python) - http://www.w3.org/2000/10/swap/doc/cwm.html
- Mailing list public-semweb-lifesci@w3.org
  - Archives: http://lists.w3.org/Archives/Public/public-semweb-lifesci/

GGF/OASIS - OGSA/OGSI/WS-RF

GGF/OASIS
Core Grid Technology Status
Effects on caBIG

Global Grid Forum

- Grid Service History
  - OGSI working group works on grid service specification
  - July 2003, OGSI 1.0 Released
  - GGF10: WS-RF announced to replace OGSI 1.0, convergence between grid and web services.
  - WSRF moved to OASIS (GGF10)

Global Grid Forum

- Data Service History
  - Data Services extend Grid Services to expose data.
  - DAIS (Data Access and Integration Services) Working Group started at GGF5
  - Specifications have evolved, current incarnation built on top of OGSI
    - Base Data Service Specification
    - XML Data Service Specification
    - Relational Database Data Service Specification
  - Possible Future Specifications
    - File Specification
    - Object Specification
  - Group is working on mapping the Data Service Specification to WSRF.

Globus

- Globus 3 (Released and Available)
  - OGSI Reference Implementation
- Globus 4
  - WSRF Reference Implementation
  - Will be Released (January 31, 2005)
  - GT4 services are not protocol-compatible with GT3 services.

OGSA-DAI

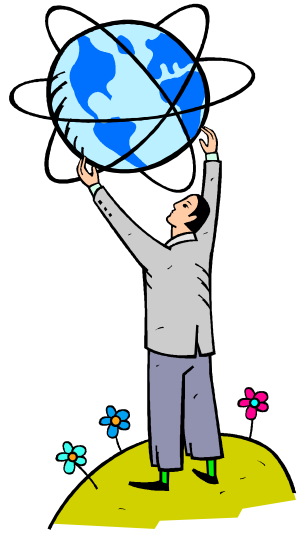
- DAIS reference Implementation
- SQL support with some XML support.
- Current release built on top of Globus 3 / OGSI
- April 2005 release built on top of Globus 4/ WSRF
- Possible Integration with Existing Grid Technologies
  - Mobius
  - Indiana University data streams
  - Datacutter
  - Storm
  - Storage Resource Broker
  - DataMiner

Technologies and caBIG

- OGSI grid technologies are more stable at this time, but will not be maintained.
- WSRF is the current standard, grid technologies will move to this, new technologies will be built on it.

Semantic Grid

1. What is Semantic Grid
2. Grid infrastructure driven by metadata.
3. Coupling Semantic Web and Grid
4. Semantic web and the grid



* Notes form Semantic Grid tutorial at GGF12.

Semantic grid in a nutshell

The Semantic Grid is an extension of the current Grid in which information and services are given well-defined and explicitly represented meaning, better enabling computers and people to work in cooperation

- Semantics in and on the Grid
- Grid with Semantics
- Intelligent Grid middleware

Grid infrastructure driven by metadata.

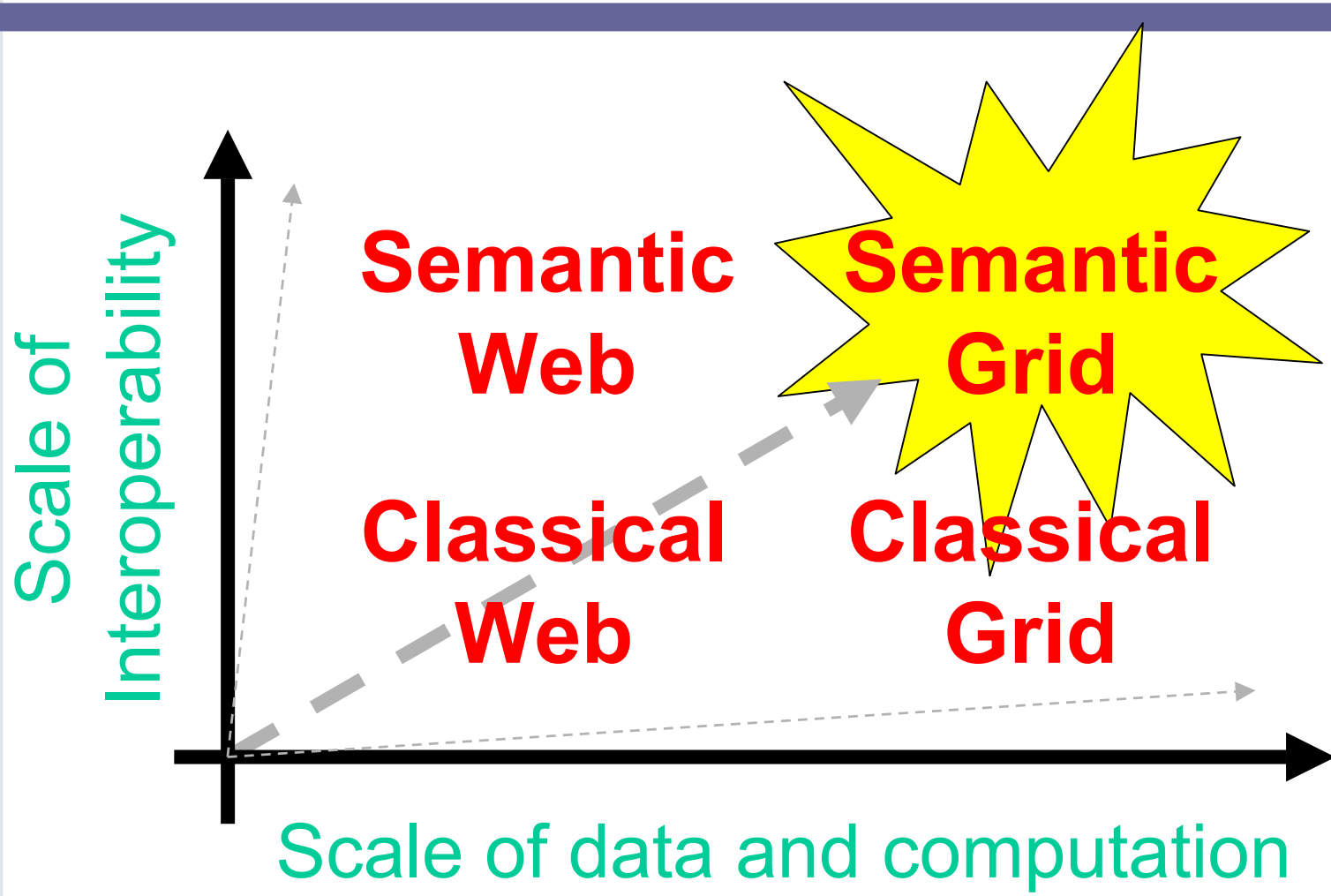
- Declarative specification of services and their requirements
- Classification of computational and data resources, performance metrics, job control; schema integration, workflow descriptions, resource brokering, resource scheduling, service state, event notification topics, typing service inputs and outputs, provenance trails; access rights to databases, personal profiles and security groupings; charging infrastructure …
- Problem solving selection and intelligent portals…

Managing and operating a Grid intelligently requires the interpretation of knowledge about the state and properties of Grid components, and their configurations for solving problems

Challenges

- Dynamic formation and management of virtual organisations
- Online negotiation of access to services: who, what, why, when, how
- Configuration of applications and systems able to deliver multiple qualities of service
- policy
- Autonomic management of distributed infrastructures, services, and applications
- Management of distributed state …

Semantic Grid



Based on an idea by Norman Paton

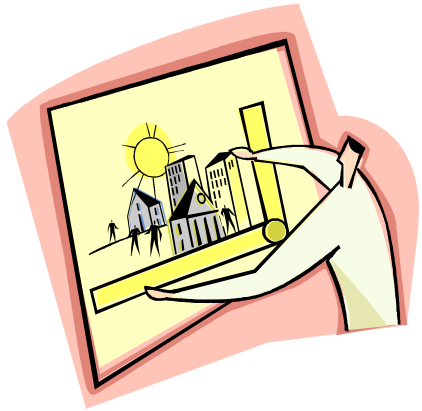
Semantic web and the grid

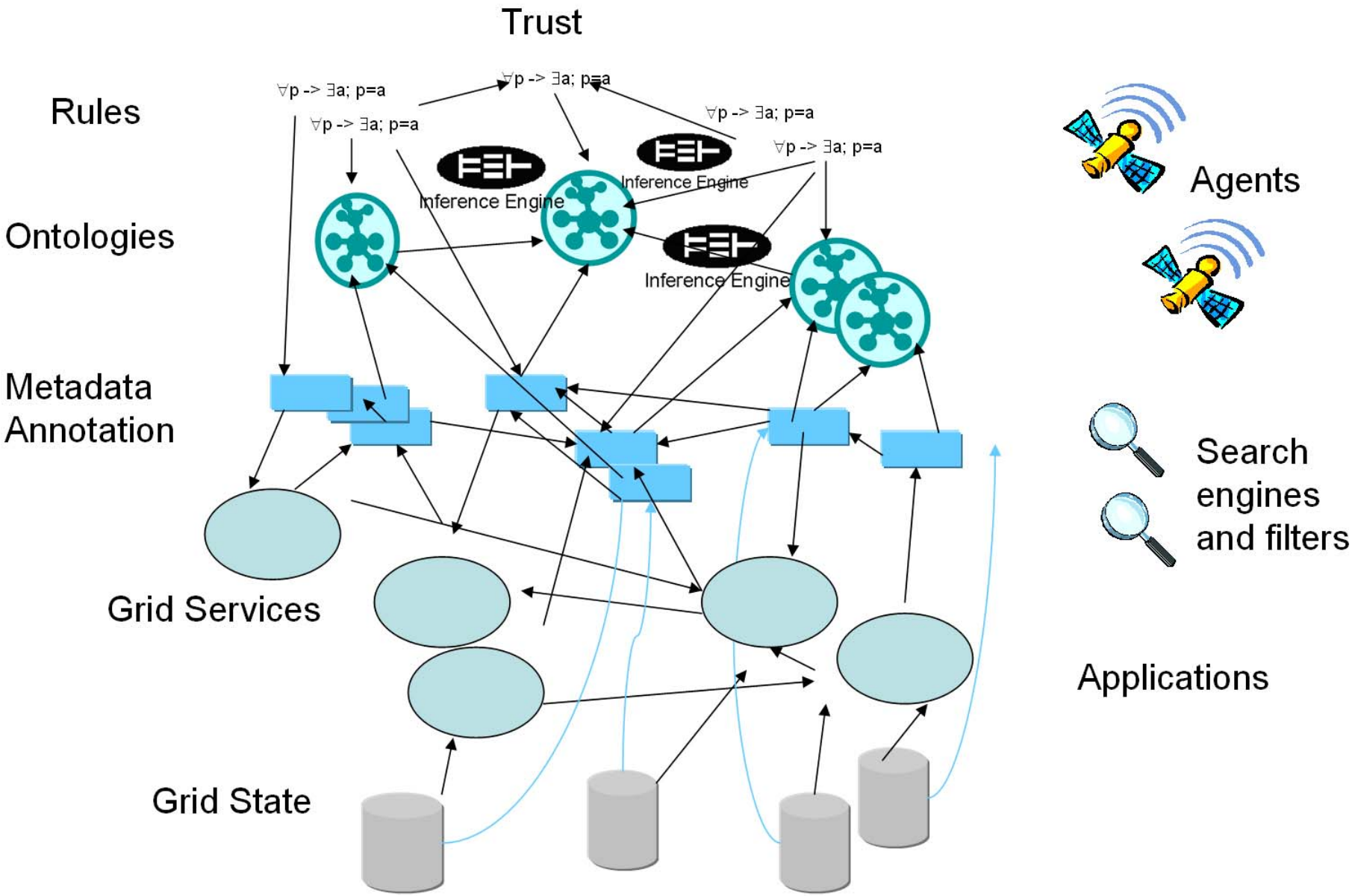
The Grid	The Semantic Web
On demand transparently constructed multi-organisational federations of distributed services	An automatically processable, machine understandable web
Distributed computing middleware	Distributed knowledge and information management
Programmatic integration, originally based on protocols & toolkits	Information integration, based on metadata, ontologies and reasoning
Information & compute power *as* a utility	Information & knowledge *is* the new utility
Application pull: pioneers are application scientists with large scale collaboration problems, originally computationally-oriented.	Technology push: pioneers are primarily from the knowledge, agent and A.I. communities.
Scalability and performance	Er, … yet to be proven

Semantic grid reference

- http://www.semanticgrid.org
- http://www.semanticgrid.org/GGF/ggf12/

3. caBIG preliminary System Components




Trust
Rules
Ontologies
Metadata
Annotation
Grid Services
Grid State
∀p -> ∃a; p=a
∀p -> ∃a; p=a
∀p -> ∃a; p=a
∀p -> ∃a; p=a
∀p -> ∃a; p=a
Inference Engine
Inference Engine
Inference Engine
Agents
Search
engines
and filters
Applications

Future actions

- Feedback from:
  - Use case producers.
  - caBIG workspaces.
  - caBIG Management team.
- Create use case specifications.
- Prioritize use cases.
- Create caBIG system architecture.
- Identify reference implementations.
- Reference implementations includes update to phase 1.
- Technology evaluation.